STATISTICS (STAT)

STAT 1010 Introductory Business Statistics
Data summaries and descriptive statistics; introduction to a statistical computer package; Probability: distributions, expectation, variance, covariance, portfolios, central limit theorem; statistical inference of univariate data; Statistical inference for bivariate data: inference for intrinsically linear simple regression models. This course will have a business focus, but is not inappropriate for students in the college. This course may be taken concurrently with the prerequisite with instructor permission.
Fall or Spring
Mutually Exclusive: STAT 1018, STAT 1110
Prerequisite: MATH 1070 OR MATH 1400 OR MATH 1100
1 Course Unit

STAT 1018 Introductory Business Statistics
The STAT 1018 honors section will cover the fundamentals of statistics through the lens of a skeptical statistician. Students will be introduced to the R language, which we will use exclusively in the course for data analysis and graphical presentations. Using examples from the upcoming election, the pandemic and other topics, we will critically examine both well-accepted and controversial claims. Examples: Can we trust election polls anymore, and are some provably more reliable than others? Should everyone get a booster shot? More generally, we will cover the basics (using a textbook costing less than $20), in order that you can use data to answer the following four questions: 1. What are the chances? 2. What's the best estimate? 3. Is there a difference? 4. How are these things related? No prior knowledge of programming, probability or statistics is required for this course.
Fall
Mutually Exclusive: STAT 1010, STAT 1110
Prerequisite: MATH 1070 OR MATH 1400 OR MATH 1100
1 Course Unit

STAT 1020 Introductory Business Statistics
Continuation of STAT 1010 or STAT 1018. A thorough treatment of multiple regression, model selection, analysis of variance, linear logistic regression; introduction to time series. Business applications. This course may be taken concurrently with the prerequisite with instructor permission.
Fall or Spring
Mutually Exclusive: STAT 1028, STAT 1120, STAT 4310
Prerequisite: STAT 1010 OR STAT 1018
1 Course Unit

STAT 1028 Introductory Business Statistics
Honors continuation of STAT 1010 or STAT 1018. A thorough treatment of multiple regression, model selection, analysis of variance, linear logistic regression; introduction to time series. Business applications. This course may be taken concurrently with the prerequisite with instructor permission.
Mutually Exclusive: STAT 1020, STAT 1120, STAT 4310
Prerequisite: STAT 1010 OR STAT 1018
1 Course Unit

STAT 1110 Introductory Statistics
Introduction to concepts in probability. Basic statistical inference procedures of estimation, confidence intervals and hypothesis testing directed towards applications in science and medicine. The use of the JMP statistical package. Knowledge of high school algebra is required for this course.
Fall or Spring
Mutually Exclusive: STAT 1010
1 Course Unit

STAT 1120 Introductory Statistics
Further development of the material in STAT 1110, in particular the analysis of variance, multiple regression, non-parametric procedures and the analysis of categorical data. Data analysis via statistical packages. This course may be taken concurrently with the prerequisite with instructor permission.
Fall or Spring
Mutually Exclusive: STAT 1020
Prerequisite: STAT 1110
1 Course Unit

STAT 3990 Independent Study
Written permission of instructor and the department course coordinator required to enroll in this course.
Fall or Spring
0.5-1 Course Unit

STAT 4010 Sports Analytics: A Capstone Course
This course would introduce undergraduate students to the growing field of sports analytics, while allowing them to implement and integrate their knowledge base by exploring real sports data sets to solve real problems. While the context will be sports related, the skills and techniques gained will be widely applicable and generalizable with applications in diverse areas. Prerequisites: Must be a declared Statistics Concentrator or Business Analytics Concentrator or Statistics Minor or Data Science Minor. Permission from the Instructor is required.
0.5 Course Units

STAT 4050 Statistical Computing with R
The goal of this course is to introduce students to the R programming language and related eco-system. This course will provide a skill-set that is in demand in both the research and business environments. In addition, R is a platform that is used and required in other advanced classes taught at Wharton, so that this class will prepare students for these higher level classes and electives.
Fall or Spring
Mutually Exclusive: STAT 7050
Prerequisite: STAT 1020 OR STAT 1120 OR STAT 4300
0.5 Course Units

STAT 4100 Data Collection and Acquisition: Strategies and Platforms
This course will give students a solid grasp of different data collection strategies and when and how they can be applied in practice. At the same time, important current ideas such as data confidentiality and ethical considerations will be addressed. The students will have designed and fielded a sample survey and designed and fielded an online experiment (A/B test). Student will collect data through web scraping activities and/or using an API. Students will summarize their collected data and subsequent inferences, culminating with an in-class presentation.
The course is structured in two parts. The first part is a "Strategies" component that addresses different data collection strategies. It will discuss sample designs, experimentation, and observational studies. The second part of the course is about "Platforms" and goes into the practicalities of the implementation of the different strategies. Given the data science perspective of this course, this is focused on web enabled approaches. Familiarity with either R or Python is expected and specifically the R-Studio or Jupyter notebooks platforms. Courses such as Stat 4050 or Stat 4770 would meet this requirement. Statistics, through the level of multiple regression is required. This requirement may be fulfilled with undergraduate courses such as Stat 1020 or Stat 1120.
Mutually Exclusive: STAT 7100
0.5 Course Units

STAT 4220 Predictive Analytics for Business
This course follows from the introductory regression classes, STAT 1020, STAT 1120, and STAT 4310 for undergraduates and STAT 6130 for MBAs. It extends the ideas from regression modeling, focusing on the core business task of predictive analytics as applied to realistic business related data sets. In particular it introduces automated model selection tools, such as stepwise regression and various current model selection criteria such as AIC and BIC. It delves into classification methodologies such as logistic regression. It also introduces classification and regression trees (CART) and the popular predictive methodology known as the random forest. By the end of the course the student will be familiar with and have applied all these tools and will be ready to use them in a work setting. The methodologies can all be implemented in either the JMP or R software packages. This course may be taken concurrently with the prerequisite with instructor permission.
Fall or Spring
Mutually Exclusive: STAT 4230, STAT 7220, STAT 7230
Prerequisite: STAT 1020 OR STAT 1120 OR STAT 4310
0.5-1 Course Unit

STAT 4230 Applied Machine Learning in Business
This course introduces students to machine learning techniques used in business applications. The main topics include: cross validation, variable selection procedures, shrinkage methods such as lasso, logistic regression, k-nearest neighbors, ROC curves and confusion matrix, trees, kernel based learning, resampling techniques, random forests, boosting, neural networks & deep learning, matrix methods including singular value decomposition (SVD) and its application in principal component analysis (PCA), and some unsupervised methods such as k-means and density based clustering. Students will learn to apply these methods in a wide range of settings such as marketing and finance, and will gain hands-on experience through class assignments and competitions. This course may be taken concurrently with the prerequisite with instructor permission.
Mutually Exclusive: STAT 4220, STAT 7220, STAT 7230
Prerequisite: STAT 1020 OR STAT 1120 OR STAT 4310
1 Course Unit

STAT 4240 Text Analytics
This course introduces modern text analytics, and the tools of natural language processing. Text and language are powerful repositories of knowledge and information, but the semi-structured nature of language makes deriving insights from text challenging. Modern analytic techniques introduced in this course make it significantly easier even for non-specialists to use text and language data to drive deep insights. The course will use several examples from real world applications in different industries such as ecommerce, healthcare and finance to illustrate these techniques. Students should be familiar with regression models at the level of Stat 6130 or Stat 1020, and the Python language at the level of Stat 4770 or Stat 7770. Familiarity with the Jupyter notebook development environment is presumed, as well as common Python packages such as pandas, NLTK and SpaCy. Those with more knowledge of Statistics, such as from Stat 7220/4220, or computing skills will benefit. The predominant software used in the course is Jupyter notebooks that use a Python interpreter. Familiarity with basic probability models is helpful but not presumed.
Fall or Spring
Mutually Exclusive: STAT 7240
0.5 Course Units

STAT 4300 Probability
Discrete and continuous sample spaces and probability; random variables, distributions, independence; expectation and generating functions; Markov chains and recurrence theory.
Fall or Spring
Mutually Exclusive: STAT 5100
Prerequisite: MATH 1080 OR MATH 1410 OR MATH 1510
1 Course Unit

STAT 4310 Statistical Inference
Graphical displays; one- and two-sample confidence intervals; one- and two-sample hypothesis tests; one- and two-way ANOVA; simple and multiple linear least-squares regression; nonlinear regression; variable selection; logistic regression; categorical data analysis; goodness-of-fit tests. A methodology course. This course does not have business applications but has significant overlap with STAT 1010 and 1020. This course may be taken concurrently with the prerequisite with instructor permission.
Fall or Spring
Mutually Exclusive: STAT 1020, STAT 5110
Prerequisite: STAT 4300
1 Course Unit

STAT 4320 Mathematical Statistics
An introduction to the mathematical theory of statistics. Estimation, with a focus on properties of sufficient statistics and maximum likelihood estimators. Hypothesis testing, with a focus on likelihood ratio tests and the consequent development of "t" tests and hypothesis tests in regression and ANOVA. Nonparametric procedures. This course may be taken concurrently with the prerequisite with instructor permission.
Spring
Mutually Exclusive: STAT 5120
Prerequisite: STAT 4300 OR STAT 5100
1 Course Unit

STAT 4330 Stochastic Processes
An introduction to Stochastic Processes. The primary focus is on Markov Chains, Martingales and Gaussian Processes. We will discuss many interesting applications from physics to economics. Topics may include: simulations of path functions, game theory and linear programming, stochastic optimization, Brownian Motion and Black-Scholes.
Fall or Spring
Mutually Exclusive: STAT 5330
Prerequisite: STAT 4300 AND (MATH 2400 OR MATH 3120 OR MATH 3140)
1 Course Unit

STAT 4350 Forecasting Methods for Management
This course provides an introduction to the wide range of techniques available for statistical modelling and forecasting of time series. Regression methods for decomposition models, trends and seasonality, spectral analysis, distributed lag models, autoregressive-moving average modeling, forecasting, exponential smoothing, and ARCH and GARCH models will be surveyed. The emphasis will be on applications, rather than technical foundations and derivations. The techniques will be studied critically, with examination of their usefulness and limitations. This course may be taken concurrently with the prerequisite with instructor permission.
Fall or Spring
Prerequisite: STAT 1020 OR STAT 1120 OR STAT 4310
1 Course Unit

STAT 4420 Introduction to Bayesian Data Analysis
The course will introduce data analysis from the Bayesian perspective to undergraduate students. We will cover important concepts in Bayesian probability modeling as well as estimation using both optimization and simulation-based strategies. Key topics covered in the course include hierarchical models, mixture models, hidden Markov models and Markov Chain Monte Carlo. A course in probability (STAT 4300 or equivalent); a course in statistical inference (STAT 1020, STAT 1120, STAT 4310 or equivalent); and experience with the statistical software R (at the level of STAT 4050 or STAT 4700) are recommended.
Spring
1 Course Unit

STAT 4510 Fundamentals of Actuarial Science I
This course is the usual entry point in the actuarial science program. It is required for students who plan to concentrate or minor in actuarial science. It can also be taken by others interested in the mathematics of personal finance and the use of mortality tables. For future actuaries, it provides the necessary knowledge of compound interest and its applications, and basic life contingencies definition to be used throughout their studies. Non-actuaries will be introduced to practical applications of finance mathematics, such as loan amortization and bond pricing, and premium calculation of typical life insurance contracts. Main topics include annuities, loans and bonds; basic principles of life contingencies and determination of annuity and insurance benefits and premiums. This course may be taken concurrently with the prerequisite with instructor permission.
Fall
Also Offered As: BEPP 4510
Prerequisite: MATH 1400 AND STAT 4300
1 Course Unit

STAT 4520 Fundamentals of Actuarial Science II
This specialized course is usually only taken by Wharton students who plan to concentrate in actuarial science and Penn students who plan to minor in actuarial mathematics. It provides a comprehensive analysis of advanced life contingencies problems such as reserving, multiple life functions, multiple decrement theory with application to the valuation of pension plans. This course may be taken concurrently with the prerequisite with instructor permission.
Spring
Also Offered As: BEPP 4520
Prerequisite: STAT 4510 OR BEPP 4510
1 Course Unit

STAT 4530 Actuarial Statistics
This course covers models for insurer's losses, and applications of Markov chains. Poisson processes, including extensions such as non-homogeneous, compound, and mixed Poisson processes are studied in detail. The compound model is then used to establish the distribution of losses. An extensive section on Markov chains provides the theory to forecast future states of the process, as well as numerous applications of Markov chains to insurance, finance, and genetics. The course is abundantly illustrated by examples from the insurance and finance literature. While most of the students taking the course are future actuaries, other students interested in applications of statistics may discover in class many fascinating applications of stochastic processes and Markov chains. This course may be taken concurrently with the prerequisite with instructor permission.
Fall
Also Offered As: BEPP 4530
Prerequisite: STAT 4300
1 Course Unit

STAT 4700 Data Analytics and Statistical Computing
This course will introduce a high-level programming language, called R, that is widely used for statistical data analysis. Using R, we will study and practice the following methodologies: data cleaning, feature extraction; web scrubbing, text analysis; data visualization; fitting statistical models; simulation of probability distributions and statistical models; statistical inference methods that use simulations (bootstrap, permutation tests). Prerequisite: Waiving the Statistics Core completely if prerequisites are not met. This course may be taken concurrently with the prerequisite with instructor permission.
Fall or Spring
Prerequisite: (STAT 1010 AND STAT 1020) OR (STAT 1110 AND STAT 1120) OR STAT 4310 OR (ECON 2300 AND ECON 2310)
1 Course Unit

STAT 4710 Modern Data Mining
With the advent of the internet age, data are being collected at unprecedented scale in almost all realms of life, including business, science, politics, and healthcare. Data mining—the automated extraction of actionable insights from data—has revolutionized each of these realms in the 21st century. The objective of the course is to teach students the core data mining skills of exploratory data analysis, selecting an appropriate statistical methodology, applying the methodology to the data, and interpreting the results. The course will cover a variety of data mining methods including linear and logistic regression, penalized regression (including lasso and ridge regression), tree-based methods (including random forests and boosting), and deep learning. Students will learn the conceptual basis of these methods as well as how to apply them to real data using the programming language R. This course may be taken concurrently with the prerequisite with instructor permission.
Fall or Spring
Prerequisite: STAT 1020 OR STAT 1120 OR STAT 4310
1 Course Unit

STAT 4740 Modern Regression for the Social, Behavioral and Biological Sciences
Function estimation and data exploration using extensions of regression analysis: smoothers, semiparametric and nonparametric regression, and supervised machine learning. Conceptual foundations are addressed as well as hands-on use for data analysis. This course may be taken concurrently with the prerequisite with instructor permission.
Spring
Also Offered As: CRIM 4740
Prerequisite: STAT 1020 OR STAT 1120
1 Course Unit

STAT 4750 Sample Survey Design
This course will cover the design and analysis of sample surveys. Topics include simple sampling, stratified sampling, cluster sampling, graphics, regression analysis using complex surveys and methods for handling nonresponse bias. This course may be taken concurrently with the prerequisite with instructor permission.
Not Offered Every Year
Prerequisite: STAT 1020 OR STAT 1120 OR STAT 4310
1 Course Unit

STAT 4760 Applied Probability Models in Marketing
This course will expose students to the theoretical and empirical "building blocks" that will allow them to construct, estimate, and interpret powerful models of consumer behavior. Over the years, researchers and practitioners have used these models for a wide variety of applications, such as new product sales, forecasting, analyses of media usage, and targeted marketing programs. Other disciplines have seen equally broad utilization of these techniques. The course will be entirely lecture-based with a strong emphasis on real-time problem solving. Most sessions will feature sophisticated numerical investigations using Microsoft Excel. Much of the material is highly technical.
Spring
Also Offered As: MKTG 4760
1 Course Unit

STAT 4770 Introduction to Python for Data Science
The goal of this course is to introduce the Python programming language within the context of the closely related areas of statistics and data science. Students will develop a solid grasp of Python programming basics, as they are exposed to the entire data science workflow, starting from interacting with SQL databases to query and retrieve data, through data wrangling, reshaping, summarizing, analyzing and ultimately reporting their results. Competency in Python is a critical skill for students interested in data science. Prerequisites: No prior programming experience is expected, but statistics, through the level of multiple regression is required. This requirement may be fulfilled with Undergraduate courses such as Stat 1020, Stat 1120.
Also Offered As: OIDD 4770
0.5-1 Course Unit

STAT 4800 Advanced Statistical Computing
This course will build on the fundamental concepts introduced in the prerequisite courses to allow students to acquire knowledge and programming skills in large-scale data analysis, data visualization, and stochastic simulation. Prerequisite: STAT 7700 or 7050 or equivalent background acquired through a combination of online courses that teach the R language and practical experience. This course may be taken concurrently with the prerequisite with instructor permission.
Spring
Prerequisite: STAT 4050 OR STAT 4700
1 Course Unit

STAT 4900 Causal Inference
Questions about cause are at the heart of many everyday decisions and public policies. Does eating an egg every day cause people to live longer or shorter or have no effect? Do gun control laws cause more or less murders or have no effect? Causal inference is the subfield of statistics that considers how we should make inferences about such questions. This course will cover the key concepts and methods of causal inference rigorously. The course is intended for statistics concentrators and minors. Knowledge of R such as that covered in STAT 4050 or STAT 4700 is recommended.
Spring
Mutually Exclusive: STAT 5900
Prerequisite: STAT 4300 AND (STAT 1020 OR STAT 1120 OR STAT 4310)
1 Course Unit

STAT 5000 Applied Regression and Analysis of Variance
An applied graduate level course in multiple regression and analysis of variance for students who have completed an undergraduate course in basic statistical methods. Emphasis is on practical methods of data analysis and their interpretation. Covers model building, general linear hypothesis, residual analysis, leverage and influence, one-way anova, two-way anova, factorial anova. Primarily for doctoral students in the managerial, behavioral, social and health sciences. Permission of instructor required to enroll.
Fall
Also Offered As: BSTA 5500, PSYC 6110
1 Course Unit

STAT 5010 Introduction to Nonparametric Methods and Log-linear Models
An applied graduate level course for students who have completed an undergraduate course in basic statistical methods. Covers two unrelated topics: loglinear and logit models for discrete data and nonparametric methods for nonnormal data. Emphasis is on practical methods of data analysis and their interpretation. Primarily for doctoral students in the managerial, behavioral, social and health sciences. Permission of instructor required to enroll.
Spring
Also Offered As: PSYC 6120
1 Course Unit

STAT 5030 Data Analytics and Statistical Computing
This course will introduce a high-level programming language, called R, that is widely used for statistical data analysis. Using R, we will study and practice the following methodologies: data cleaning, feature extraction; web scrubbing, text analysis; data visualization; fitting statistical models; simulation of probability distributions and statistical models; statistical inference methods that use simulations (bootstrap, permutation tests).
Prerequisite: Two courses at the statistics 4000 or 5000 level.
Fall or Spring
1 Course Unit

STAT 5100 Probability
Elements of matrix algebra. Discrete and continuous random variables and their distributions. Moments and moment generating functions. Joint distributions. Functions and transformations of random variables. Law of large numbers and the central limit theorem. Point estimation: sufficiency, maximum likelihood, minimum variance. Confidence intervals. A one-year course in calculus is recommended.
Fall or Spring
Mutually Exclusive: STAT 4300
1 Course Unit

STAT 5110 Statistical Inference
Graphical displays; one- and two-sample confidence intervals; one- and two-sample hypothesis tests; one- and two-way ANOVA; simple and multiple linear least-squares regression; nonlinear regression; variable selection; logistic regression; categorical data analysis; goodness-of-fit tests. A methodology course.
Fall or Spring
Mutually Exclusive: STAT 1020, STAT 4310
Prerequisite: STAT 5100
1 Course Unit

STAT 5120 Mathematical Statistics
An introduction to the mathematical theory of statistics. Estimation, with a focus on properties of sufficient statistics and maximum likelihood estimators. Hypothesis testing, with a focus on likelihood ratio tests and the consequent development of "t" tests and hypothesis tests in regression and ANOVA. Nonparametric procedures.
Spring
Mutually Exclusive: STAT 4320
Prerequisite: STAT 4300 OR STAT 5100
1 Course Unit

STAT 5150 Advanced Statistical Inference I
STAT 5150 is aimed at first-year Ph.D. students and builds a good foundation in statistical inference from the first principles of probability.
Fall
Prerequisite: STAT 4300 AND STAT 4310 AND MATH 2400
1 Course Unit

STAT 5160 Advanced Statistical Inference II
STAT 5160 is a natural continuation of STAT 5150, and the main focus is on asymptotic evaluations and regression models. Time permitting, it also discusses some basic nonparametric statistical methods.
Spring
Prerequisite: STAT 5150
1 Course Unit

STAT 5200 Applied Econometrics I
This is a course in econometrics for graduate students. The goal is to prepare students for empirical research by studying econometric methodology and its theoretical foundations. Students taking the course should be familiar with elementary statistical methodology and basic linear algebra, and should have some programming experience. Topics include conditional expectation and linear projection, asymptotic statistical theory, ordinary least squares estimation, the bootstrap and jackknife, instrumental variables and two-stage least squares, specification tests, systems of equations, generalized least squares, and introduction to use of linear panel data models.
Fall
Prerequisite: (MATH 1080 OR MATH 1410) AND MATH 3120
1 Course Unit

STAT 5210 Applied Econometrics II
Topics include system estimation with instrumental variables, fixed effects and random effects estimation, M-estimation, nonlinear regression, quantile regression, maximum likelihood estimation, generalized method of moments estimation, minimum distance estimation, and binary and multinomial response models. Both theory and applications will be stressed.
Spring
Prerequisite: STAT 5200
1 Course Unit

STAT 5330 Stochastic Processes
An introduction to Stochastic Processes. The primary focus is on Markov Chains, Martingales and Gaussian Processes. We will discuss many interesting applications from physics to economics. Topics may include: simulations of path functions, game theory and linear programming, stochastic optimization, Brownian Motion and Black-Scholes.
Fall or Spring
Mutually Exclusive: STAT 4330
Prerequisite: STAT 5100
1 Course Unit

STAT 5350 Forecasting Methods for Management
This course provides an introduction to the wide range of techniques available for statistical modelling and forecasting of time series. Regression methods for decomposition models, trends and seasonality, spectral analysis, distributed lag models, autoregressive-moving average modeling, forecasting, exponential smoothing, and ARCH and GARCH models will be surveyed. The emphasis will be on applications, rather than technical foundations and derivations. The techniques will be studied critically, with examination of their usefulness and limitations.
Fall or Spring
1 Course Unit

STAT 5420 Bayesian Methods and Computation
Sophisticated tools for probability modeling and data analysis from the Bayesian perspective. Hierarchical models, mixture models and Monte Carlo simulation techniques.
Spring
Prerequisite: STAT 4300 OR STAT 5100
1 Course Unit

STAT 5710 Modern Data Mining
Modern Data Mining: Statistics or Data Science has been evolving rapidly to keep up with the modern world. While classical multiple regression and logistic regression technique continue to be the major tools we go beyond to include methods built on top of linear models such as LASSO and Ridge regression. Contemporary methods such as KNN (K nearest neighbor), Random Forest, Support Vector Machines, Principal Component Analyses (PCA), the bootstrap and others are also covered. Text mining especially through PCA is another topic of the course. While learning all the techniques, we keep in mind that our goal is to tackle real problems. Not only do we go through a large collection of interesting, challenging real-life data sets but we also learn how to use the free, powerful software "R" in connection with each of the methods exposed in the class. Prerequisite: two courses at the statistics 4000 or 5000 level or permission from instructor.
Fall or Spring
1 Course Unit

STAT 5800 Advanced Statistical Computing
This course will build on the fundamental concepts introduced in the prerequisite courses to allow students to acquire knowledge and programming skills in large-scale data analysis, data visualization, and stochastic simulation. Prerequisite: STAT 5030, 7050, or 7700 or equivalent background acquired through a combination of online courses that teach the R language and practical experience.
Spring
Prerequisite: STAT 5030 OR STAT 7050 OR STAT 7700
1 Course Unit

STAT 5900 Causal Inference
Questions about cause are at the heart of many everyday decisions and public policies. Does eating an egg every day cause people to live longer or shorter or have no effect? Do gun control laws cause more or less murders or have no effect? Causal inference is the subfield of statistics that considers how we should make inferences about such questions. This course will cover the key concepts and methods of causal inference rigorously. Background in probability and statistics; some knowledge of R is recommended.
Spring
Mutually Exclusive: STAT 4900
1 Course Unit

STAT 6130 Regression Analysis for Business
This course provides the fundamental methods of statistical analysis, the art and science if extracting information from data. The course will begin with a focus on the basic elements of exploratory data analysis, probability theory and statistical inference. With this as a foundation, it will proceed to explore the use of the key statistical methodology known as regression analysis for solving business problems, such as the prediction of future sales and the response of the market to price changes. The use of regression diagnostics and various graphical displays supplement the basic numerical summaries and provides insight into the validity of the models. Specific important topics covered include least squares estimation, residuals and outliers, tests and confidence intervals, correlation and autocorrelation, collinearity, and randomization. The presentation relies upon computer software for most of the needed calculations, and the resulting style focuses on construction of models, interpretation of results, and critical evaluation of assumptions.
Fall
Prerequisite: STAT 6110
1 Course Unit

STAT 6210 Accelerated Regression Analysis for Business
STAT 6210 is intended for students with recent, practical knowledge of the use of regression analysis in the context of business applications. This course covers the material of STAT 6130, but omits the foundations to focus on regression modeling. The course reviews statistical hypothesis testing and confidence intervals for the sake of standardizing terminology and introducing software, and then moves into regression modeling. The pace presumes recent exposure to both the theory and practice of regression and will not be accommodating to students who have not seen or used these methods previously. The interpretation of regression models within the context of applications will be stressed, presuming knowledge of the underlying assumptions and derivations. The scope of regression modeling that is covered includes multiple regression analysis with categorical effects, regression diagnostic procedures, interactions, and time series structure. The presentation of the course relies on computer software that will be introduced in the initial lectures. Recent exposure to the theory and practice of regression modeling is recommended.
Fall
0.5 Course Units

STAT 7010 Modern Data Mining
Modern Data Mining: Statistics or Data Science has been evolving rapidly to keep up with the modern world. While classical multiple regression and logistic regression technique continue to be the major tools we go beyond to include methods built on top of linear models such as LASSO and Ridge regression. Contemporary methods such as KNN (K nearest neighbor), Random Forest, Support Vector Machines, Principal Component Analyses (PCA), the bootstrap and others are also covered. Text mining especially through PCA is another topic of the course. While learning all the techniques, we keep in mind that our goal is to tackle real problems. Not only do we go through a large collection of interesting, challenging real-life data sets but we also learn how to use the free, powerful software "R" in connection with each of the methods exposed in the class. Prerequisite: two courses at the statistics 4000 or 5000 level or permission from instructor.
Fall or Spring
1 Course Unit

STAT 7050 Statistical Computing with R
The goal of this course is to introduce students to the R programming language and related eco-system. This course will provide a skill-set that is in demand in both the research and business environments. In addition, R is a platform that is used and required in other advanced classes taught at Wharton, so that this class will prepare students for these higher level classes and electives.
Fall or Spring
Mutually Exclusive: STAT 4050
Prerequisite: STAT 6130 OR STAT 6210
0.5 Course Units

STAT 7100 Data Collection and Acquisition: Strategies and Platforms
This course will give students a solid grasp of different data collection strategies and when and how they can be applied in practice. At the same time, important current ideas such as data confidentiality and ethical considerations will be addressed. The students will have designed and fielded a sample survey and designed and fielded an online experiment (A/B test). Student will collect data through web scraping activities and/or using an API. Students will summarize their collected data and subsequent inferences, culminating with an in-class presentation.
The course is structured in two parts. The first part is a "Strategies" component that addresses different data collection strategies. It will discuss sample designs, experimentation, and observational studies. The second part of the course is about "Platforms" and goes into the practicalities of the implementation of the different strategies. Given the data science perspective of this course, this is focused on web enabled approaches. Familiarity with either R or Python is expected and specifically the R-Studio or Jupyter notebooks platforms. Courses such as Stat 7050 or Stat 7770 would meet this requirement. Statistics, through the level of multiple regression is required. This requirement may be fulfilled with MBA courses such as Stat 6130/6210, or by waiving MBA statistics.
Mutually Exclusive: STAT 4100
0.5 Course Units

STAT 7110 Forecasting Methods for Management
This course provides an introduction to the wide range of techniques available for statistical modelling and forecasting of time series. Regression methods for decomposition models, trends and seasonality, spectral analysis, distributed lag models, autoregressive-moving average modeling, forecasting, exponential smoothing, and ARCH and GARCH models will be surveyed. The emphasis will be on applications, rather than technical foundations and derivations. The techniques will be studied critically, with examination of their usefulness and limitations. This course may be taken concurrently with the prerequisite with instructor permission.
Fall or Spring
Prerequisite: STAT 6130 OR STAT 6210
1 Course Unit

STAT 7220 Predictive Analytics for Business
This course follows from the introductory regression classes, STAT 1020, STAT 1120, and STAT 4310 for undergraduates and STAT 6130 for MBAs. It extends the ideas from regression modeling, focusing on the core business task of predictive analytics as applied to realistic business related data sets. In particular it introduces automated model selection tools, such as stepwise regression and various current model selection criteria such as AIC and BIC. It delves into classification methodologies such as logistic regression. It also introduces classification and regression trees (CART) and the popular predictive methodology known as the random forest. By the end of the course the student will be familiar with and have applied all these tools and will be ready to use them in a work setting. The methodologies can all be implemented in either the JMP or R software packages. This course is formerly STAT 6220.
Fall or Spring
Mutually Exclusive: STAT 4220, STAT 4230, STAT 7230
Prerequisite: STAT 6130 OR STAT 6210
0.5-1 Course Unit

STAT 7230 Applied Machine Learning in Business
This course introduces students to machine learning techniques used in business applications. The main topics include: cross validation, variable selection procedures, shrinkage methods such as lasso, logistic regression, k-nearest neighbors, ROC curves and confusion matrix, trees, kernel based learning, resampling techniques, random forests, boosting, neural networks & deep learning, matrix methods including singular value decomposition (SVD) and its application in principal component analysis (PCA), and some unsupervised methods such as k-means and density based clustering. Students will learn to apply these methods in a wide range of settings such as marketing and finance, and will gain hands-on experience through class assignments and competitions.
Mutually Exclusive: STAT 4220, STAT 4230, STAT 7220
Prerequisite: STAT 6130 OR STAT 6210
1 Course Unit

STAT 7240 Text Analytics
This course introduces modern text analytics, and the tools of natural language processing. Text and language are powerful repositories of knowledge and information, but the semi-structured nature of language makes deriving insights from text challenging. Modern analytic techniques introduced in this course make it significantly easier even for non-specialists to use text and language data to drive deep insights. The course will use several examples from real world applications in different industries such as ecommerce, healthcare and finance to illustrate these techniques. Students should be familiar with regression models at the level of Stat 6130 or Stat 1020, and the Python language at the level of Stat 4770 or Stat 7770. Familiarity with the Jupyter notebook development environment is presumed, as well as common Python packages such as pandas, NLTK and SpaCy. Those with more knowledge of Statistics, such as from Stat 7220/4220, or computing skills will benefit. The predominant software used in the course is Jupyter notebooks that use a Python interpreter. Familiarity with basic probability models is helpful but not presumed.
Fall or Spring
Mutually Exclusive: STAT 4240
0.5 Course Units

STAT 7700 Data Analytics and Statistical Computing
This course will introduce a high-level programming language, called R, that is widely used for statistical data analysis. Using R, we will study and practice the following methodologies: data cleaning, feature extraction; web scrubbing, text analysis; data visualization; fitting statistical models; simulation of probability distributions and statistical models; statistical inference methods that use simulations (bootstrap, permutation tests).
Prerequisite: Two courses at the statistics 4000 or 5000 level.
Fall or Spring
1 Course Unit

STAT 7760 Applied Probability Models in Marketing
This course will expose students to the theoretical and empirical "building blocks" that will allow them to construct, estimate, and interpret powerful models of consumer behavior. Over the years, researchers and practitioners have used these models for a wide variety of applications, such as new product sales, forecasting, analyses of media usage, and targeted marketing programs. Other disciplines have seen equally broad utilization of these techniques. The course will be entirely lecture-based with a strong emphasis on real-time problem solving. Most sessions will feature sophisticated numerical investigations using Microsoft Excel. Much of the material is highly technical.
Spring
Also Offered As: MKTG 7760
1 Course Unit

STAT 7770 Introduction to Python for Data Science
The goal of this course is to introduce the Python programming language within the context of the closely related areas of statistics and data science. Students will develop a solid grasp of Python programming basics, as they are exposed to the entire data science workflow, starting from interacting with SQL databases to query and retrieve data, through data wrangling, reshaping, summarizing, analyzing and ultimately reporting their results. Competency in Python is a critical skill for students interested in data science. Prerequisites: No prior programming experience is expected, but statistics, through the level of multiple regression is required. This requirement may be fulfilled with MBA courses such as STAT 6130/6210; or by waiving MBA statistics.
Also Offered As: OIDD 7770
0.5-1 Course Unit

STAT 7800 Advanced Statistical Computing
This course will build on the fundamental concepts introduced in the prerequisite courses to allow students to acquire knowledge and programming skills in large-scale data analysis, data visualization, and stochastic simulation. Prerequisite: STAT 5030, 7050, or 7700 or equivalent background acquired through a combination of online courses that teach the R language and practical experience.
Spring
Prerequisite: STAT 5030 OR STAT 7050 OR STAT 7700
1 Course Unit

STAT 8510 Fundamentals of Actuarial Science I
This course is the usual entry point in the actuarial science program. It is required for students who plan to concentrate or minor in actuarial science. It can also be taken by others interested in the mathematics of personal finance and the use of mortality tables. For future actuaries, it provides the necessary knowledge of compound interest and its applications, and basic life contingencies definition to be used throughout their studies. Non-actuaries will be introduced to practical applications of finance mathematics, such as loan amortization and bond pricing, and premium calculation of typical life insurance contracts. Main topics include annuities, loans and bonds; basic principles of life contingencies and determination of annuity and insurance benefits and premiums. Prerequisite: One semester of calculus.
Fall
Also Offered As: BEPP 8510
1 Course Unit

STAT 8520 Fundamentals of Actuarial Science II
This specialized course is usually only taken by Wharton students who plan to concentrate in actuarial science and Penn students who plan to minor in actuarial mathematics. It provides a comprehensive analysis of advanced life contingencies problems such as reserving, multiple life functions, multiple decrement theory with application to the valuation of pension plans.
Spring
Also Offered As: BEPP 8520
Prerequisite: STAT 8510 OR BEPP 8510
1 Course Unit

STAT 8530 Actuarial Statistics
This course covers models for insurer's losses, and applications of Markov chains. Poisson processes, including extensions such as non-homogeneous, compound, and mixed Poisson processes are studied in detail. The compound model is then used to establish the distribution of losses. An extensive section on Markov chains provides the theory to forecast future states of the process, as well as numerous applications of Markov chains to insurance, finance, and genetics. The course is abundantly illustrated by examples from the insurance and finance literature. While most of the students taking the course are future actuaries, other students interested in applications of statistics may discover in class many fascinating applications of stochastic processes and Markov chains. Prerequisite: Two semesters of statistics.
Fall
Also Offered As: BEPP 8530
1 Course Unit

STAT 8990 Independent Study
Written permission of instructor, the department MBA advisor and course coordinator required to enroll.
Fall or Spring
0.5-1 Course Unit

STAT 9150 Nonparametric Inference
Statistical inference when the functional form of the distribution is not specified. Nonparametric function estimation, density estimation, survival analysis, contingency tables, association, and efficiency.
Not Offered Every Year
Prerequisite: STAT 5200
1 Course Unit

STAT 9200 Sample Survey Methods
This course will cover the design and analysis of sample surveys. Topics include simple random sampling, stratified sampling, cluster sampling, graphics, regression analysis using complex surveys and methods for handling nonresponse bias.
Not Offered Every Year
Prerequisite: STAT 5200 OR STAT 9610 OR STAT 9700
1 Course Unit

STAT 9210 Observational Studies
This course will cover statistical methods for the design and analysis of observational studies. Topics will include the potential outcomes framework for causal inference; randomized experiments; matching and propensity score methods for controlling confounding in observational studies; tests of hidden bias; sensitivity analysis; and instrumental variables.
Fall or Spring
Prerequisite: STAT 5200 OR STAT 9610 OR STAT 9700
1 Course Unit

STAT 9250 Multivariate Analysis: Theory
This is a course that prepares PhD students in statistics for research in multivariate statistics and high dimensional statistical inference. Topics from classical multivariate statistics include the multivariate normal distribution and the Wishart distribution; estimation and hypothesis testing of mean vectors and covariance matrices; principal component analysis, canonical correlation analysis and discriminant analysis; etc. Topics from modern multivariate statistics include the Marcenko-Pastur law, the Tracy-Widom law, nonparametric estimation and hypothesis testing of high-dimensional covariance matrices, high-dimensional principal component analysis, etc.
Not Offered Every Year
Prerequisite: STAT 9300 OR STAT 9700 OR STAT 9720
1 Course Unit

STAT 9260 Multivariate Analysis: Methodology
This is a course that prepares PhD students in statistics for research in multivariate statistics and data visualization. The emphasis will be on a deep conceptual understanding of multivariate methods to the point where students will propose variations and extensions to existing methods or whole new approaches to problems previously solved by classical methods. Topics include: principal component analysis, canonical correlation analysis, generalized canonical analysis; nonlinear extensions of multivariate methods based on optimal transformations of quantitative variables and optimal scaling of categorical variables; shrinkage- and sparsity-based extensions to classical methods; clustering methods of the k-means and hierarchical varieties; multidimensional scaling, graph drawing, and manifold estimation.
Not Offered Every Year
Prerequisite: STAT 9610
1 Course Unit

STAT 9270 Bayesian Statistical Theory and Methods
This graduate course will cover the modeling and computation required to perform advanced data analysis from the Bayesian perspective. We will cover fundamental topics in Bayesian probability modeling and implementation, including recent advances in both optimization and simulation-based estimation strategies. Key topics covered in the course include hierarchical and mixture models, Markov Chain Monte Carlo, hidden Markov and dynamic linear models, tree models, Gaussian processes and nonparametric Bayesian strategies.
Not Offered Every Year
Prerequisite: STAT 4300 OR STAT 5100
1 Course Unit

STAT 9280 Statistical Learning Theory
Statistical learning theory studies the statistical aspects of machine learning and automated reasoning, through the use of (sampled) data. In particular, the focus is on characterizing the generalization ability of learning algorithms in terms of how well they perform on "new" data when trained on some given data set. The focus of the course is on: providing the fundamental tools used in this analysis; understanding the performance of widely used learning algorithms; understanding the "art" of designing good algorithms, both in terms of statistical and computational properties. Potential topics include: empirical process theory; online learning; stochastic optimization; margin based algorithms; feature selection; concentration of measure. Background in probability and linear algebra recommended.
Spring
1 Course Unit

STAT 9300 Probability Theory
Measure theoretic foundations, laws of large numbers, large deviations, distributional limit theorems, Poisson processes, random walks, stopping times.
Fall
Also Offered As: AMCS 6481, MATH 6480
Prerequisite: STAT 4300 OR STAT 5100 OR MATH 6080
1 Course Unit

STAT 9310 Stochastic Processes
Continuation of MATH 6480/STAT 9300, the 2nd part of Probability Theory for PhD students in the math or statistics department. The main topics include Brownian motion, martingales, Ito's formula, and their applications to random walk and PDE.
Not Offered Every Year
Also Offered As: AMCS 6491, MATH 6490
1 Course Unit

STAT 9550 Stochastic Calculus and Financial Applications
Selected topics in the theory of probability and stochastic processes.
Fall
Prerequisite: STAT 9300
1 Course Unit

STAT 9600 Statistical Algorithms and Computation
This course aims to prepare students for graduate work in the design, analysis, and implementation of statistical algorithms. The target audience is Ph.D. students in statistics or in adjacent fields, such as computer science, mathematics, electrical engineering, computational biology, economics, and marketing. We will take a fundamental approach and focus on classes of algorithms of primary importance in statistics and statistical machine learning. Some meta-classes of algorithms that may receive significant attention are optimization, sampling, and numerical linear algebra. I aim to make the content complementary rather than overlapping with other courses at Penn, such as ESE6050, CIS6770, and the CIS7000 series. While there may be some overlap in the portions of the course that cover optimization, the sampling (Monte Carlo and related) aspects of the course are, to my knowledge, hard to find elsewhere at Penn. The course is fast paced and I expect a certain degree of mathematical preparation. Most students in the above mentioned programs will have the requisite mathematics background. I also expect familiarity with an appropriate programming language such as R, python, or matlab. The course will be mostly language agnostic. However, I may at times give example code in one of these languages, and you will be expected to be able to read the code even if it is not in your "primary" language. We may make use of various open-source toolboxes and packages for these environments, such as the Stan probabilistic programming language (best used with R) and the cvx toolbox for convex programming (available for multiple platforms but perhaps best used with matlab).
Spring
1 Course Unit

STAT 9610 Statistical Methodology
This is a course that prepares 1st year PhD students in statistics for a research career. This is not an applied statistics course. Topics covered include: linear models and their high-dimensional geometry, statistical inference illustrated with linear models, diagnostics for linear models, bootstrap and permutation inference, principal component analysis, smoothing and cross-validation.
Fall
Prerequisite: STAT 4310 OR STAT 5200
1 Course Unit

STAT 9620 Advanced Methods for Applied Statistics
This course is designed for Ph.D. students in statistics and will cover various advanced methods and models that are useful in applied statistics. Topics for the course will include missing data, measurement error, nonlinear and generalized linear regression models, survival analysis, experimental design, longitudinal studies, building R packages and reproducible research.
Spring
Prerequisite: STAT 9610
1 Course Unit

STAT 9700 Mathematical Statistics
Decision theory and statistical optimality criteria, sufficiency, point estimation and hypothesis testing methods and theory.
Fall
Prerequisite: STAT 4310 OR STAT 5200
1 Course Unit

STAT 9710 Introduction to Linear Statistical Models
Theory of the Gaussian Linear Model, with applications to illustrate and complement the theory. Distribution theory of standard tests and estimates in multiple regression and ANOVA models. Model selection and its consequences. Random effects, Bayes, empirical Bayes and minimax estimation for such models. Generalized (Log-linear) models for specific non-Gaussian settings.
Spring
Prerequisite: STAT 9700
1 Course Unit

STAT 9720 Advanced Topics in Mathematical Statistics
A continuation of STAT 9700.
Fall or Spring
Prerequisite: STAT 9700 AND STAT 9710
1 Course Unit

STAT 9740 Modern Regression for the Social, Behavioral and Biological Sciences
Function estimation and data exploration using extensions of regression analysis: smoothers, semiparametric and nonparametric regression, and supervised machine learning. Conceptual foundations are addressed as well as hands-on use for data analysis.
Spring
Prerequisite: STAT 1020 OR STAT 1120
1 Course Unit

STAT 9910 Seminar in Advanced Application of Statistics
This seminar will be taken by doctoral candidates after the completion of most of their coursework. Topics vary from year to year and are chosen from advance probability, statistical inference, robust methods, and decision theory with principal emphasis on applications.
Fall or Spring
0.5-1 Course Unit

STAT 9950 Dissertation
Fall or Spring
0 Course Units

STAT 9990 Independent Study
Written permission of instructor and the department course coordinator required to enroll.
Fall or Spring
0.5-2 Course Units